Edge-Centric Multimodal Authentication System Using Encrypted Biometric Templates

[1]Zulfiqar Ali, [2,3,4]M. Shamim Hossain, [1,4]Ghulam Muhammad, [5]Ihsan Ullah, [4]Hamid Abachi, [2,3,4]Atif Alamri

[1]Digital Speech Processing Group, College of Computer and Information Sciences
King Saud University, Riyadh 11543, Saudi Arabia
[2]Research Chair of Pervasive and Mobile Computing, College of Computer and Information Sciences, King Saud University, Riyadh 11543, Saudi Arabia
[3]Department of Software Engineering, College of Computer and Information Sciences
King Saud University, Riyadh 11543, Saudi Arabia
[4]Department of Computer Engineering, College of Computer and Information Sciences
King Saud University, Riyadh 11543, Saudi Arabia
[5]Insight Centre for Data Analytics, National University of Ireland
Galway, Ireland

Corresponding Author: mshossain@ksu.edu.sa

ABSTRACT

Data security, complete system control, and missed storage and computing opportunities in personal portable devices are some of the major limitations of the centralized cloud environment. Among these limitations, security is a prime concern due to potential unauthorized access to private data. Biometrics, in particular, is considered sensitive data, and its usage is subject to the privacy protection law. To address this issue, a multimodal authentication system using encrypted biometrics for the edge-centric cloud environment is proposed in this study. Personal portable devices are utilized for encrypting biometrics in the proposed system, which optimizes the use of resources and tackles another limitation of the cloud environment. Biometrics is encrypted using a new method. In the proposed system, the edges transmit the encrypted speech and face for processing in the cloud. The cloud then decrypts the biometrics and performs authentication to confirm the identity of an individual. The model for speech authentication is based on two types of features, namely, Mel-frequency cepstral coefficients and perceptual linear prediction coefficients. The model for face authentication is implemented by determining the eigenfaces. The final decision about the identity of a user is based on majority voting. Experimental results show that the new encryption method can reliably hide the identity of an individual and accurately decrypt the biometrics, which is vital for errorless authentication.

INDEX TERMS: cloud computing, privacy protection, biometric templates, encryption, chaotic system, ORL database.

1. Introduction

The recent development in communication technologies enables the use of computing as a utility. The deployment of high-capacity hardware at the user-facing end is no longer required to run applications and Internet services that need storage, computation, and communication. The cloud can provide the platform for such services and applications by supplying centralized resources in a reliable and cost-effective manner. Moreover, cloud analytics is utilized to analyze industry data by using a range of analytical tools and methods. The process data can then be used to draw conclusions and make decisions. According to the prediction of Gartner, approximately 90% of deployed data will be useless by 2018. Therefore, transmitting only relevant data collected by the IoT to the cloud for analysis is important [1, 2]. Deploying the edges of low-cost, low-capacity, and low-performance devices in the designed architecture of an application or service for cloud computing is possible as well. In this manner, the data can be filtered through some intelligent methods, and the features of some deployed services and applications can be enhanced. For instance, with the help of edge analytics, the sensors deployed for traffic monitoring can also be used to send an alert to the fire brigade in case of fire by analyzing the surroundings.

Loss of security in transmitting personal and social data is one of the fundamental problems in cloud computing [3-6]. Numerous smart healthcare systems have been developed to monitor patients in smart homes and cities [7, 8]. In these smart systems, the patient's data collected by the IoT are transmitted to health centers for diagnosis. Unauthorized access to such sensitive data may create unavoidable circumstances in someone's personal and professional life. For instance, telemedicine has been successfully applied in various areas of medical fields [9-11]. In all of its types [12], that is, store and send, self-monitoring, and interactive, the patient's data travel through wireless communication to specialists and consultants. If patient data are vulnerable in such applications, such as a cosmetic surgery breach[1], then the patients may face embarrassments and humiliations. Similarly, biometric authentication is required in some applications, such as remote access to secret data and bank accounts [13]. Therefore, individual data transmitted through wireless channels should be secured to avoid financial losses and job termination[14, 15]. The main objective of the present study is to develop a secure biometric authentication system based on encrypted biometric templates.

The identity of a person can be verified in the biometric authentication systems using personal attributes, such as speech [16, 17], face [18, 19], fingerprints [20, 21], palmprint [22, 23], gait [24, 25], and iris [26, 27]. These physiological and behavioral attributes of humans are more reliable in authentication compared with knowledge-based or token-based approaches because these attributes cannot be stolen and are unique for every individual. However, the authentication system based on a single biometric is sometimes unable to recognize a person correctly. Therefore, various multimodal authentication systems based on more than one biometrics have been developed for the accurate authentication of a person. In [28], a multimodal biometric system based on speech, face, and fingerprint is proposed. The system simultaneously uses all biometrics to make the final decision about the identity of a person. Although the authors claim that the multimodal authentication system overcomes the limitations of a single biometric, a comparison of each biometric against the multimodal authentication system is not provided. Therefore, the improvement in the

[1]https://www.theguardian.com/technology/2017/may/31/hackers-publish-private-photos-cosmetic-surgery-clinic-bitcoin-ransom-payments

case of the multimodal authentication system cannot be noted. In another study [29], a multimodal authentication system based on speech, face, and fingerprint is analyzed for threshold adjustment to ensure that it performs better than the unimodal biometric system. Then, the computed thresholds are used in score level fusion to reach a final decision about the identity of a user. Ribaric et al. also developed a bimodal verification system using palmprint and face [30]. Their experimental results showed that the performance of the bimodal system is close to the results of palmprint authentication. However, significant improvement is observed when the system is compared with face verification. Overall, improvements of 0.74% and 1.72% are attained for the equal error rate (EER) and total error rate, respectively. A multimodal biometric system based on fingerprint and iris recognition is developed in [31]. For authentication, a person is recognized using the fingerprints and iris, and the final decision about the identity is obtained by performing the AND operation between outputs of fingerprint and iris recognition. The authors did not provide the accuracy for each biometric, and the improvement in the case of bimodal authentication is not mentioned.

Several multimodal biometric systems have been developed using speech as one of the biometrics. Kartik et al. developed a bimodal biometric authentication system using speech and signature [32]. Verification of a speaker is done by extracting the Mel-frequency cepstral coefficients (MFCC) [33] and using vector quantization for pattern matching [33]. The maximum obtained accuracies for clean and noisy speech are 100% and 73.75%, respectively. Meanwhile, the highest accuracies for signature recognition with clean and noisy data are 80% and 72.92%, respectively. In the case of bimodal authentication, the system shows an improvement of 1.25% for noisy data. The authors extended their work in [34], and three biometrics, namely, face, speech, and signatures, are considered to implement the biometric authentication system. In the developed system, a 6% improvement is reported for multimodal authentications, while the accuracies of unimodal authentications using face, speech, and fingerprint are 82.5%, 86.67%, and 92.92%, respectively. Kumar et al. also developed a multimodal biometric system using speech and face images. For speaker verification, MFCC with Gaussian mixture model (GMM) is implemented and an EER of 8% is obtained. Face recognition is conducted using principal component analysis (PCA) and linear discriminant analysis. The obtained EER is 22.87%, an improvement of 1% compared with that in unimodal authentication.

According to the European Union General Data Protection Regulation 2016/679, the biometrics of individuals are sensitive data whose use is protected under privacy protection rights [35]. The biometric template must be protected to avoid any leakage of sensitive data [36]. A number of multi-biometric template protection approaches are listed in [37], yet no method for multimodal biometrics system in the encryption domain has been developed so far. Although a general framework that uses homomorphic encryption for the protection of templates in the multimodal authentication system is proposed in [37], the designed framework is developed for fingerprint and signature verification.

In the present study, an encrypted multimodal biometric system based on speech and face images is proposed. The proposed system authenticates a user remotely through the cloud environment. The biometrics travel via wireless communication for processing, which is risky and makes the data vulnerable. To avoid risks, a new method for biometric encryption is suggested and implemented in the proposed system. Moreover, to optimize computing resources, the biometrics are encrypted by the edges, which is another concern in cloud computing in addition to security [3]. Through the encryption, the transmitted biometrics will not be in the original form but in the encrypted form instead. Therefore, the biometrics will

not be exposed to threats of data breach. To investigate the new method of encryption in the proposed multimodal authentication system, a user is verified by using the original and encrypted speech and face. The model for speech authentication is developed using two well-known speech features, MFCC and perceptual linear prediction coefficient (PLP), both of which are used with GMM. By contrast, the model for face recognition is based on eigenfaces and Euclidean distance (EUD). The experimental results indicate that the user identity is properly hidden after the encryption and cannot be disclosed unless it is decrypted. Several experiments are likewise performed to ensure that the new encryption method recovered the identity accurately. The results show no observable difference between the original and decrypted signals. The contributions of this study are summarized as follows:

- A new encryption and decryption method for multi-biometric template protection.
- Optimization of computing resources in the cloud environment by encrypting the biometrics through edges.
- Biometric decryption in the cloud, and double authentication of a user using speech and face.
- Two approaches of verification for speech and one approach for face recognition. The final decision is based on majority voting.

The remainder of this paper is organized as follows. Section 2 describes the proposed biometrics protected authentication system. Section 3 elaborates the new method of encryption to encrypt the face and speech signals of a user and determines the reliability of the encryption and decryption processes. Section 4 provides the authentication results of the proposed system using the original and encrypted faces and signals. Experiments with the recovered data are also performed to show that the encryption method recovers the face and signals like the original. Finally, Section 5 draws the conclusions.

2. Proposed Biometrics Protected Authentication System

A secured multimodal biometric authentication system using encrypted templates is proposed in this study. The block diagram of the proposed multimodal authentication system is shown in Fig. 1. The system performs the encryption and decryption of the biometrics before authentication. For the encryption and decryption of the biometrics, a new method is suggested and implemented in the proposed multimodal authentication system. The proposed biometric authentication system is designed in such a manner that the biometrics collected from the IoT should not be transmitted in the original form because of risks to the privacy of the user's biometrics. Unauthorized access to the biometrics can be used illegally to take control of applications/services without the knowledge of the user. Therefore, the biometrics of the user captured through IoT are encrypted by the edges, which are low-cost devices with normal computing power. Then, encrypted data with secret key are transmitted to the cloud via wireless communication. The cloud is responsible for the decryption of the biometrics and performs the authentication to grant access to the services.

The authentication process in the proposed system does not rely on a single biometric. A bimodal authentication process using two human attributes, namely, face image and speech signals, is implemented. Moreover, the authentication by speech is performed with two types of speech features. Each type of feature provides an independent decision. However, the final decision about the identity of a user is determined

based on majority voting. This factor ensures the proposed system is robust and reliable, while the encryption of the biometrics ensure that the system is secured.

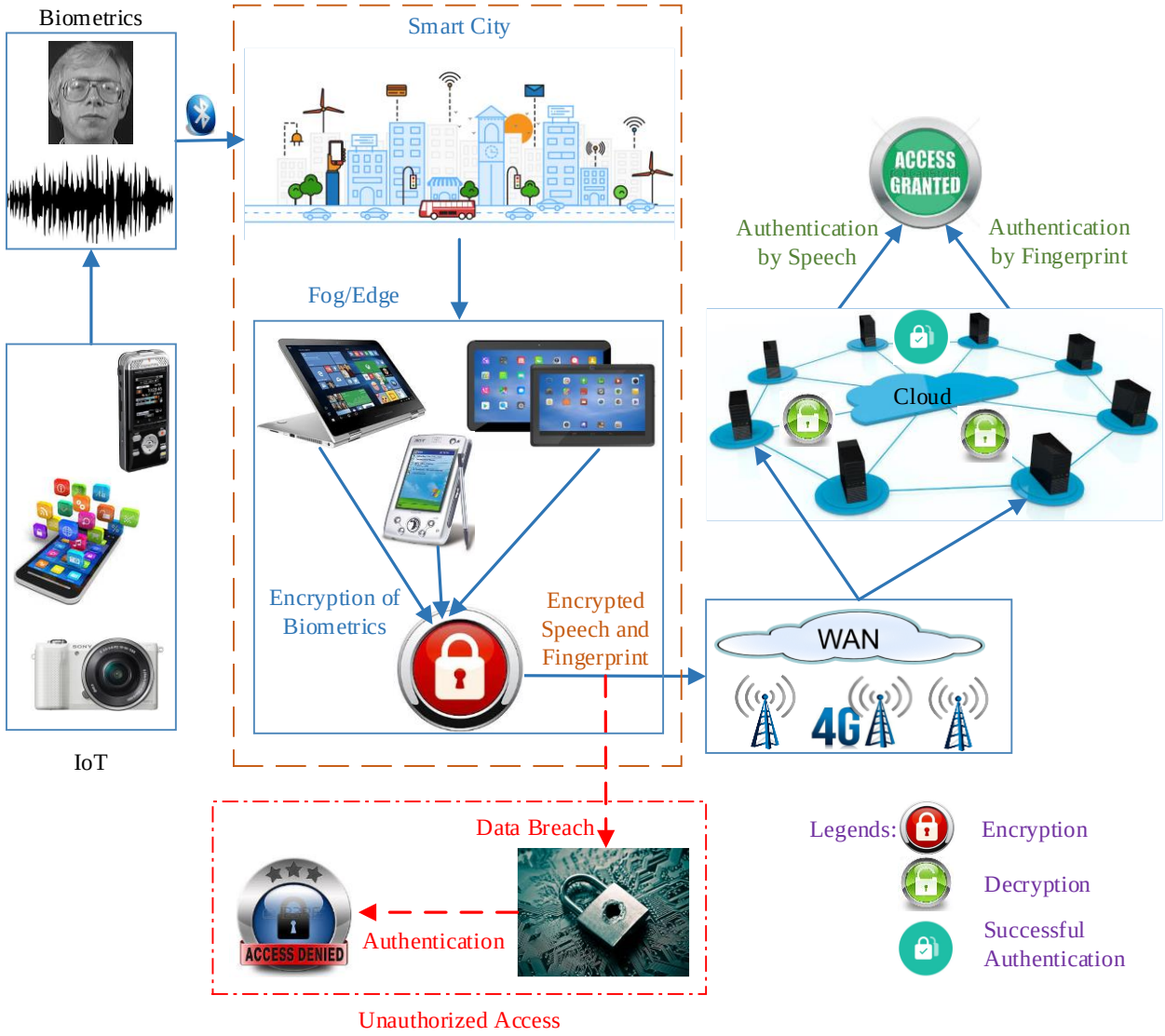


Figure 1: Proposed system for secured multimodal biometric authentication.

The implementation of bimodal authentication requires one database for face images and another database for speech signals. The face and speech are not from the same users. A chimeric database in which the face and speech of different users are assigned to one chimeric identity is used in this study to evaluate the proposed system. A chimeric database instead of a real multimodal database is also used in [38] to evaluate the multi-biometric template protection schemes. The speech signals are randomly associated with the faces, as listed in Appendix A. The face images are taken from the Database of Faces [39], which was developed between 1992 and 1994 and was formerly known as the ORL Database of Faces. This database contains 400 images of 40 distinct subjects. Each subject has 10 different face images taken by varying the lighting, facial details (wearing glasses/without glasses), and facial expressions, such as open/closed eyes and smiling/not smiling. All images are grayscale with pixel values of 0–255. The dimension of each image

is 92 by 112 pixels. All images are taken against a dark homogeneous background and stored in Portable Gray Map (PGM) format. A total of 40 directories are created to organize the database, with each directory containing all the images of a subject.

The speech signals are taken from a database recorded at the Massachusetts Eye and Ear Infirmary (MEEI) Voice and Speech Laboratory [40-43]. The database contains the speech signals of normal persons and dysphonic patients. However, the current study only considers the speech signals of normal persons. In the Database of Faces, the number of distinct faces is 40. Therefore, the same number of speech signals is drawn from the MEEI database. The normal persons in the speech database are recorded at a 25 kHz sampling frequency using a bit rate of 16 bits in a controlled environment. One reason for using the MEEI database is that it has speech signals with a long duration. Each normal person has recorded a speech of 12 s. Such long speech signals provide sufficient space for embedding the face image into it. These signals can be divided into two equal halves as well. By doing this, the model for speech authentication can be trained using the first half and the testing can be done with the second half. In this manner, we can show that the model is independent of text. This type of authentication is referred to as text-independent verification. The speech authentication model is also evaluated using text-dependent verification. The first step in the proposed system is the encryption of face images and speech signals using the new method proposed in this study. The method is described in the following subsections.

3. Encryption and Decryption Processes to Protect Biometrics

Speech signals and face images are encrypted using a new method to prevent unauthorized access to the biometrics of a user. The method has two major processes, referred to as encryption and decryption processes. During the encryption process, the method generates two chaotic sequences to randomize the biometrics so that, even in the case of data breach, nobody will be able to access the services, such as gaining control to secret data or access to bank accounts. Moreover, the proposed system has the capability to deny access by not authenticating the breached or attacked biometrics.

3.1. Encryption of Face Image and Speech Signal

The following steps describe the encryption process of the newly proposed method to encrypt the face image and speech signal of users.

1. Read face image I of user X. The dimensions of image I is I_r by I_c.
2. Read speech signal S of user X. The number of samples (N) in the signal S is equal to the product of the duration of the signal in seconds (T) and the sampling frequency (fs), i.e., $N = T \times fs$, where $\times$ represents the multiplication.
3. To encrypt face image I, generate random sequence Q^1 of the length $L = I_r \times I_c$ using the chaotic system [44, 45] given in Eq. 1 with the initial conditions λ and $Q^1{}_0$:

$$Q^1_{q+1} = \lambda Q^1_q \left(1 - Q^1_q\right). \qquad (1)$$

4. Scale the generated sequence Q^1 within the specific range $[a_1\ a_2]$. The scaled sequence sQ^1 is obtained as follows:

$$sQ^1 = a_1 + \left(\frac{Q^1 - \min(Q^1)}{\max(Q^1) - \min(Q^1)} \right) \times (a_2 - a_1) \quad . \quad (2)$$

The obtained sequence sQ^1 is reshaped such that the numbers of rows and columns are equal to I_r and I_c, respectively.

5. Transform image I by dividing its pixel values by the corresponding elements of the sequence sQ^1. The transformed image I^t is computed using Eq. 3:

$$I^t_{(i,j)} = \frac{I_{(i,j)}}{sQ^1_{(i,j)}}, \text{ where } i = 1,2,3,\ldots,I_r \text{ and } j = 1,2,3,\ldots,I_c \quad . \quad (3)$$

6. The encrypted face image EI is obtained by adding the transformed image I^t to the sequence sQ^1, as follows:

$$EI_{(i,j)} = I^t_{(i,j)} + sQ^1_{(i,j)} \quad . \quad (4)$$

7. The speech signal S is encrypted using the transformed image I^t and a random sequence Q^2 of 0's and 1's having lengths equal to N. The indices of the 1's in the sequence Q^2 are generated by the chaotic system (given in Eq. 1) with the same initial conditions λ and $Q^2{}_0$ (=$Q^1{}_0$) but scaled in a different range [a_3 a_4] using Eq. 2. The number of 1's in Q^2 is equal to the number of pixels in the transformed image, that is, L, and the number of 0's in Q^2 becomes $N - L$.
8. The encryption of the signal S is conducted using the transformed image I^t according to the following criteria:

$$(ES)_k = \begin{cases} S_k + I^t_k & \text{if } (Q^2_k = 1 \text{ and } S_k > 0) \\ S_k - I^t_k & \text{if } (Q^2_k = 1 \text{ and } S_k \le 0) \end{cases}$$

where

$k = 1,2,3,\ldots,N$. (5)

The encrypted speech signal ES and face image EI with the initial conditions (λ and $Q^1{}_0$) and the limits of the intervals [a_1 a_2 a_3 a_4] are transmitted through wireless communication to the cloud. The cloud is responsible for decrypting the received biometrics and will perform the authentication. The secret key to decrypting the biometric consists of the initial conditions and limits of intervals.

3.2. Decryption of Speech Signal and Face Image

The decryption process of the proposed method to decrypt the encrypted signal ES and face image EI using the secret key containing the initial conditions and limits of intervals is performed with the following steps:

1. Generate the chaotic sequence $Q^{1'}$ using the conditions λ and $Q^1{}_0$, and scale the sequence in the range [a_1 a_2]. The scaled sequence $sQ^{1'}$ is reshaped such that the numbers of rows and columns are equivalent to eI_r and eI_c, which represent the rows and columns of the encrypted image EI, respectively.
2. The first step to decrypt the face image EI is to obtain the transformed image $I^{t'}$ using Eq. 6:

$$I^{t'}_{(i,j)} = EI_{(i,j)} - sQ^{1'}_{(i,j)}$$

where . (6)

$i = 1,2,3,\ldots,eI_r$ and $j = 1,2,3,\ldots,eI_c$

3. Then, decrypted image I' is obtained by multiplying the corresponding values of the sequence $sQ^{1'}$ and the recovered transformed image $I^{t'}$ using the relationship expressed in Eq. 7:

$$I'_{(i,j)} = I^{t'}_{(i,j)} \times sQ^{1'}_{(i,j)} . \quad (7)$$

4. To decrypt the signal *ES*, the sequence $Q^{2'}$ is generated equal to the length of *ES*, i.e., *N*. The sequence contains only 0's and 1's, where the indices for 1's are obtained from the chaotic system (given in Eq. 1) with the conditions λ and $Q^1{}_0$ and limits [a_3 a_4]. The number of 1's in $Q^{2'}$ is *L*, while the number of 0's is $N - L$.
5. The decrypted signal S' is recovered using the recovered transformed image $I^{t'}$ and the generated sequence $Q^{2'}$ with the relationship expressed in Eq. 8:

$$S'_k = \begin{cases} (ES)_k - I^{t'}_k & \text{if} \quad (Q^{2'}_k = 1 \quad \text{and} \quad (ES)_k > 0) \\ (ES)_k + I^{t'}_k & \text{if} \quad (Q^{2'}_k = 1 \quad \text{and} \quad (ES)_k \le 0) \end{cases}$$

where

$k = 1, 2, 3, \ldots, N$. (8)

The decrypted face image I' and speech signal S' will be used by the proposed system to authenticate the identity of a user. Access will be granted if the user is genuine. Otherwise, access will be denied.

3.3. Analysis of Encryption and Decryption

In the proposed multimodal biometric authentication system, data protection and accurate biometric authentication are crucial and not subject to compromise. Analyzing the encryption process is important to ensure that the biometrics are protected and, in the case of unauthorized access, cannot be used for user authentication. The analysis of the decryption process is also conducted to evaluate the quality of the recovered biometrics and determine how close it is to the original biometrics. Furthermore, biometric authentication using the original, encrypted, and recovered biometrics is presented in Section 4.

3.3.1. Encryption Reliability

The encrypted face image *EI* and speech signal *ES* are compared with the original image *I* and signal *S* using different metrics to analyze the reliability of the encryption process. The metrics used to analyze the encryption of the face image are bit error rate (BRT), mean squared error (MSE), and peak signal-to-noise ratio (PSNR), which are expressed in Eqs. 9, 10, and 11, respectively:

$$BRT(I, EI) = \frac{ERB}{(I_r \times I_c)} \times 100 , \quad (9)$$

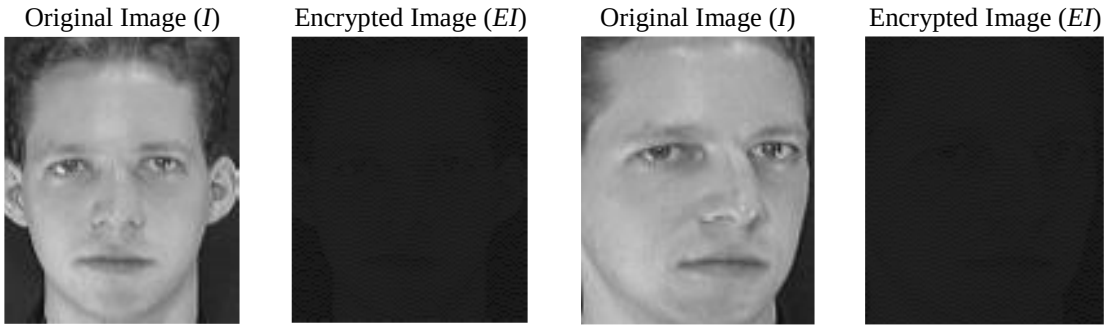
$$MSE(I, EI) = \frac{\sum_{i=1}^{I_r} \sum_{j=1}^{I_c} \left(I_{(i,j)} - EI_{(i,j)} \right)^2}{I_r \times I_c} , \quad (10)$$

$$PSNR(I, EI) = 20 \log_{10} \left(\frac{2^{BP} - 1}{MSE} \right). \quad (11)$$

In Eq. 9, ERB represents the number of erroneous bits, which denote the differences in the original and encrypted images at corresponding locations. A high BRT indicates a large difference between original and

encrypted images, which is good because the encrypted image should be significantly different from the original image. Similarly, a high MSE indicates that the original and encrypted images are different from each other. By contrast, a high PSNR denotes that the images are similar to each other, whereas a low PSNR denotes that the images are dissimilar. In Eq. 11, BP represents the number of bits per pixel in the image.

Two images of the same user are taken from the database to analyze the encryption process. To encrypt the images, the values of the initial conditions λ and Q^1_0 are 3.5 and 0.5, respectively. For these values of initial conditions, Eq. 1 generates deterministic random numbers that can be regenerated using the same initial conditions during the decryption process [46]. The generated random number are also scaled in the interval $[a_1\ a_2]$ using Eq. 2 with $a_1 = 10$ and $a_2 = 30$. The encrypted images with the computed metrics are shown in Fig. 2. The BRT and MSE between the first original image and its encrypted image are 48.93% and 21,510, respectively, which are high and indicate that the images are significantly different from each other. A low PSNR of 7.15 concludes the same thing. The first original image and its encrypted image are shown in Fig. 2(a). A similar trend is observed in the metrics of the second original image and its encrypted image. The BRT and MSE are high, whereas the PSNR is low. The second original image and its encrypted image are depicted in Fig. 2(b).



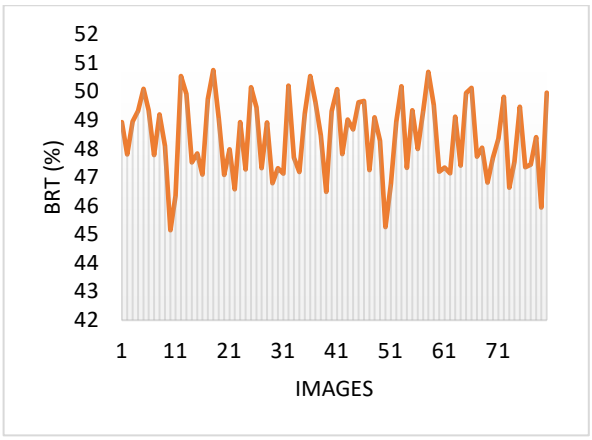
BRT = 48.93%; MSE = 21,510; PSNR = 7.15 (a)

BRT = 50.09%; MSE = 15,458; PSNR = 6.23 (b)

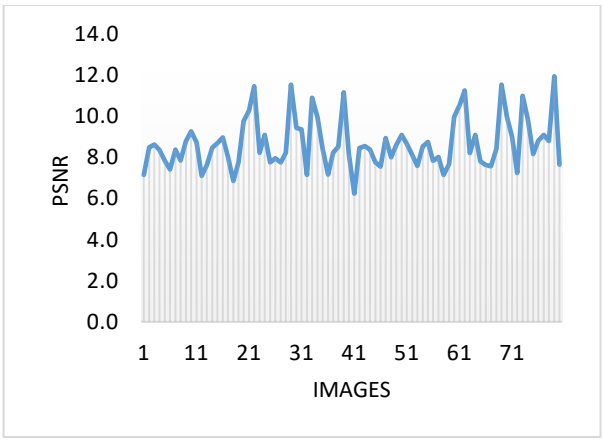
Figure 2: Encryption of images and computed metrics: (a) first image of user 1 and (b) second image of user 1.

The encrypted images shown in Figs. 2(a) and 2(b) do not disclose any information about the user. Therefore, in the case of unauthorized access, access via biometric authentication is not granted. Objective analysis of biometric authentication using encrypted images will be discussed in the next section.

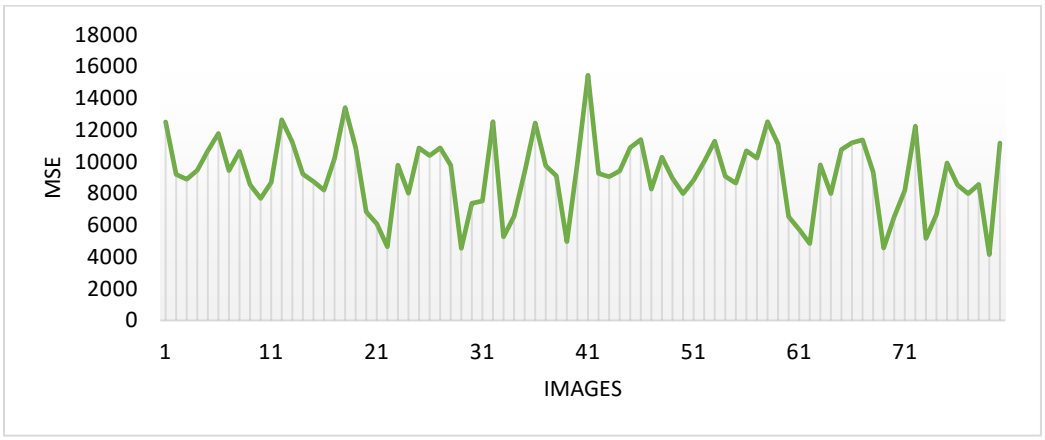
The metrics are computed for the images of all users in the database to analyze the reliablity of the encryption process. The face database contains 10 images for each user. The measures are computed using the first 2 images of all 40 users. The computed BRT, PSNR, and MSE for the encrypted measures are shown in Figs. 3(a) to 3(c), respectively. Notably, BRT for all users is more than 45%, which indicates that the encryption process is reliable and will not reveal the information of a user. Moreover, the PSNR is under 12 dB for all users, indicating that the ecrypted images are distorted and cannot be used for authentication in case of data breach. Large values of MSE likewise show that the original and encrypted images are different for all users.



(a)



(b)



(c)

Figure 3: Computed metrics for the face images of all users: (a) BRT, (b) PSNR, and (c) MSE.

The analysis of the encryption of speech signals is conducted using the EUD to determine the distortion between original and encrypted speech signals:

$$EUD(S,S') = \sqrt{\sum_{i=1}^{N}(S_i - S_i')^2}\ . \quad (12)$$

The speech signal of each user is divided into two halves. The duration of each speech signal in the MEEI database is 12 s. Therefore, each part of the signal contains a 6 s recording . Each part provides sufficient space for embedding the transformed image for signal encryption. Two parts of the original and encrypted signals of user 1 are depicted in Fig. 4. Parts 1 and 2 of the original signal and its corresponding encrypted signal are shown in Figs. 4(a) and 4(b), respectively. The encrypted signals of both parts are significantly different from the original signals. Moreover, the amplitude of the original signals lies between −1 and 1, whereas the amplitude of the encrypted signals is in the range of −20 to 20.

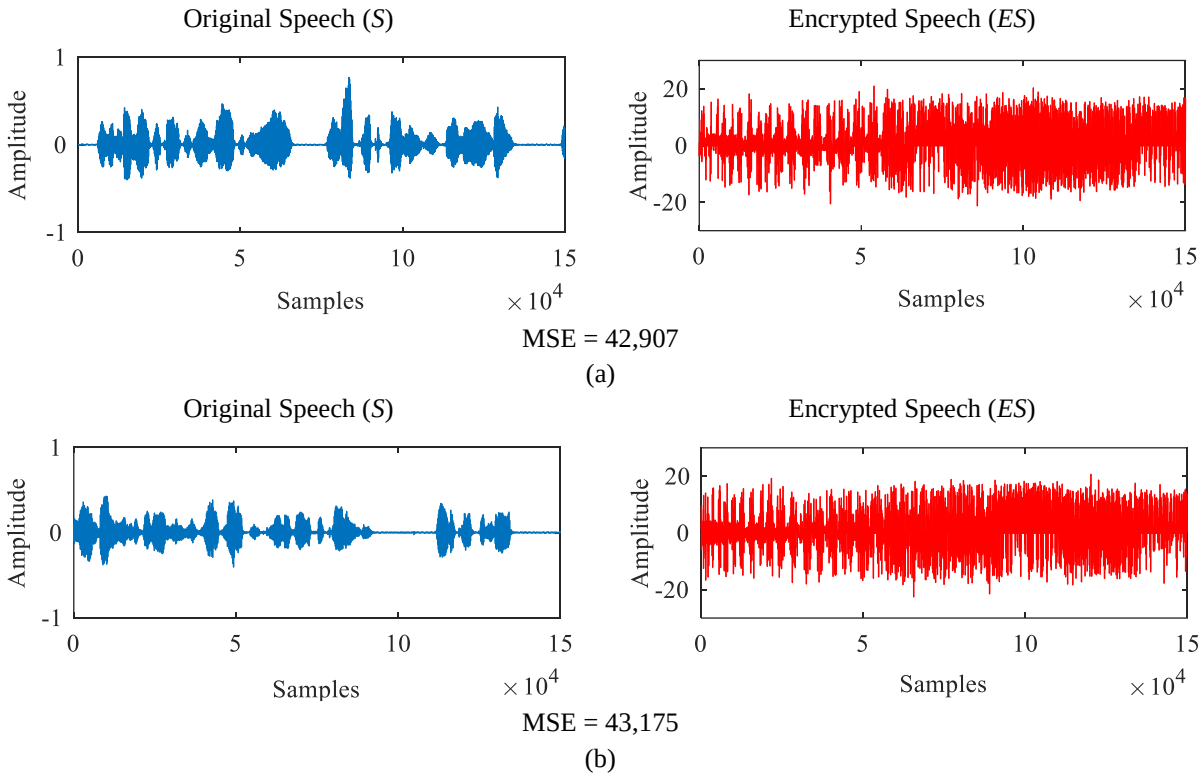


Figure 4: Encryption of speech and computed MSE: (a) the first half of the signal for user 1 and (b) the second half of the signal for user 1.

The computed EUD values for both parts of the signal are 745.4 and 825.1. The large values of MSE also confirm that the encrypted signal is significantly different from the original signal and will not allow anyone to be authenticated in case of unauthorized data access. Similarly, the speech signals of all users are encrypted, and the computed EUD is plotted in Fig. 5. The EUD is greater than 500 for all users. This large EUD indicates that all encrypted signals are distorted and different from the original signals. Objective analysis of the encryption process for biometric authenticity using the encrypted signal is presented in Section 4.

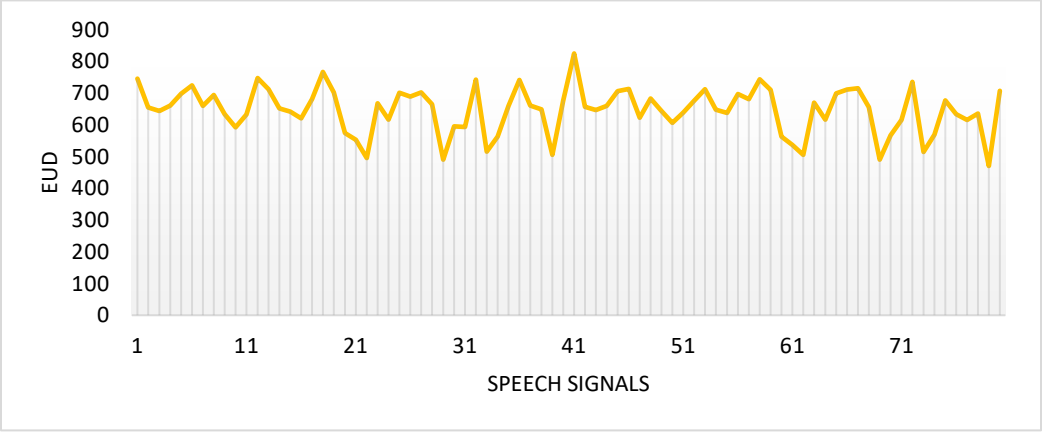


Figure 5: Computed EUD for the speech signals of all users

The analysis of encrypted face images and speech signals indicate that the encryption process of the proposed method is reliable. Therefore, the biometrics are secured after encryption, and no information can be retrieved without successful decryption of the images and signals through the relevant secret key.

3.3.2. Decryption Reliability

To authenticate the biometrics, they should be recovered like the original, which is impossible without accurate decryption. Therefore, a reliable decryption process is an essential component for an accurate authentication system based on encrypted biometrics. The encrypted images and signals are decrypted using the relevant secret key to investigate the method. The metrics BRT, MSE, and PSNR are also calculated to determine the similarity between original and decrypted images and faces.
Two decrypted face images of user 1 are shown in Figs. 6(a) and 6(b). The decrypted images are similar to the original images as confirmed by the computed metrics. The BRT, MSE, and PSNR are 0, 0, and infinity (Inf), respectively, for both decrypted faces.

Recovered Image (I')

BRT = 0; MSE = 0; PSNR = Inf

(a)

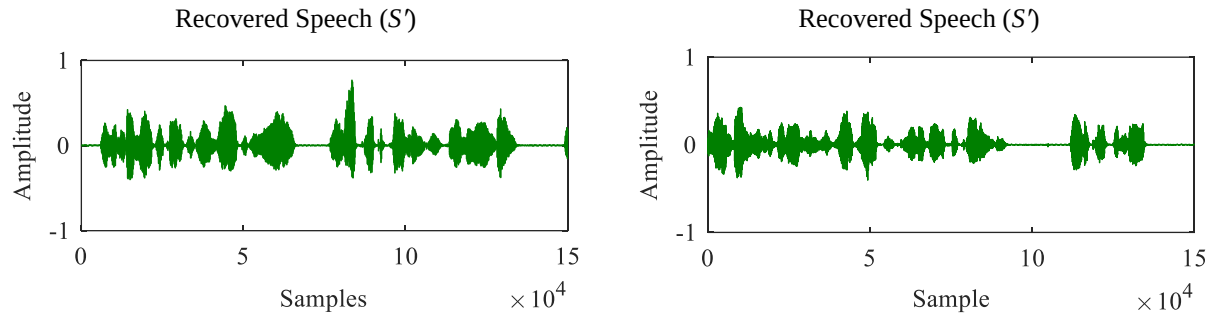
Recovered Image (I')

BRT = 0; MSE = 0; PSNR = Inf

(b)

Figure 6: Decryption of faces and computed metrics: (a) the first image of user 1 and (b) the second image of user 1.

The decrypted speech signals of user 1 are depicted in Figs. 7(a) and 7(b). Fig. 7(a) shows the decrypted signal for the first part of the signal, while Fig. 7(b) shows the decrypted signal for the second part of the signal. Both decrypted parts are equivalent to the original signal. The calculated EUD values for parts 1 and 2 of the decrypted signal are 2.3e-26 and 1.8e-26, respectively. The EUD is approximately equal to zero, which confirms that the decrypted signals are not distorted and are similar to the original signal.



EUD = 2.3622 e-26 (a)

EUD = 1.8507e-26 (b)

Figure 7: Decryption of speech signals and computed EUD: (a) the first part of user 1 and (b) the second part of user 1.

Similarly, when decryption of the image and signal of every user is performed, the decrypted images and signals are similar to the original images and signals. The decrypted images and signals do not exhibit degradation. The measures BRT, MSE, and EUD are zero, and PSNR is Inf for all users of the databases. Hence, the decryption process is reliable and capable of retrieving the images and faces in the original form. The decrypted image and signal of a user will be used to authenticate his or her identity in the proposed system.

4. Authentication of Biometrics for Access

Two models are developed to authenticate a user. The first model performs the authentication using the speech signal, whereas the second model performs the authentication using the face images.

4.1. Model for Speech Authentication

The model for speech authentication is developed using two types of speech features, namely, MFCC and PLP. In the speech authentication model, the acoustic models of every user are generated during the training phase using different numbers of Gaussian mixtures. During the testing phase, user verification is done by applying GMM.

The MFCC features mimic the human auditory system and have been used successfully in many speech-related applications, including speaker recognition. MFCC is based on one of the psychoacoustic principles of the human auditory system, which is known as critical bandwidth phenomena. The important steps in MFCC computation are segmentation of the speech signal into short frames, application of the Hamming window on each frame, computation of the short-term spectrum using Fourier transformation, filtration of the spectrum through bandpass filters spaced over Mel's scale, and finally, the use of discrete cosine transformation for decorrelation of the coefficients. The obtained coefficients are referred to as the MFCC.

The other speech feature, PLP, is compared with MFCC. PLP is based on three psychoacoustic principles of the human auditory system: critical bandwidth phenomena, equal loudness hearing curve, and intensity loudness power law of hearing [47]. These principles are applied sequentially on the power spectrum of the

input speech signal obtained from the application of Fourier transformation. Then, linear prediction (LP) analysis of order p is applied with the autocorrelation method using inverse Fourier transformation. Finally, the PLP features c_n are obtained by providing the output of the last step to the recursive relationship expressed in Eq. 13:

$$c_1 = \ln \sigma^2$$
$$c_n = a_n + \sum_{k=1}^{n-1} \left(\frac{k}{n} \right) c_k \; a_{n-k} \;, \quad 1 \le n \le p \;, \qquad (13)$$

where σ^2 represents the gain during LP analysis.

In the developed speech model, the parameters of GMM are initialized with the *k*-means algorithm [48] and tuned by applying the well-known expectation–maximization algorithm [49] to generate the model that provided the maximum log-likelihood value. To recognize an unknown user, the features are extracted from the speech signal and compared with the acoustic model of each user. One log-likelihood score for every speaker is computed, and the user having the maximum log-likelihood score with the unknown user will be the recognized user. The speech signal is divided into the short frames before feature extraction. Therefore, the final log-likelihood score for the unknown user is computed by adding the log-likelihood score of all frames.

4.2. Model for Face Authentication

Face authentication is performed by developing a face recognition model in the proposed system. The model computes the eigenfaces to recognize a user. The idea of eigenfaces worked exactly like Fourier transformation, which can be used to reconstruct a signal from the sum of weighted sinusoids perfectly. Similarly, the face of a user can be reconstructed by the sum of weighted eigenfaces.

The face images of all users are converted into a vector to compute the eigenfaces. The face images are grayscale, and the values of the pixels are in the range of 0–255. The dimension of each image is I_r by I_c, and after reshaping into a vector, the dimension becomes 1 by L, where $L = I_r \times I_c$. After the 2-D image is converted into a vector, each face image is represented by a point in the face space. Given that every user has multiple images, we can assume that the images of a particular user will lie close to one another in the face space, whereas the images of different users will be separated by a significant distance. Working with the vectors of high dimensions is also not ideal. The dimensionality should be reduced, but the resultant face space should maximize the distance between different faces. To reduce the dimensionality, PCA is applied to obtain vectors with low dimensionality in the face space. During PCA, the obtained eigenvectors γ of all registered users are sorted in descending order, and the first 20 eigenvectors are considered in this study to represent the face space.

During recognition of a user, the unknown face I_u will be projected to the face space using Eq. 14:

$$\Upsilon_u = \gamma \left(I_u - \Delta I \right)$$
where . (14)
$$\Delta I = \frac{1}{T} \sum_{t=1}^{T} I_t$$

In Eq. 14, ΔI is the average face of all registered users, γ denotes the eigenvectors of all users computed during PCA, T represents the total number of registered users, and I_t is the image vector of the user. The obtained weights of the user I_u are denoted by Υ_u. The weight Υ_u will be compared with the weights of all users, i.e., Υ_1, Υ_2, Υ_3, …, Υ_T, by computing the EUD between them to recognize user I_u. The user with the minimum distance will be recognized as an unknown user.

4.3. Experimental Results for Biometric Authentication

4.3.1. Authentication Results for Speech

The speech signals of every user are divided into two parts before encryption. Each part of the signal contains different recorded texts and provides the platform for the text-independent speaker recognition. In this type of speaker authentication, the text used to authenticate the user is different from that used to train the model. Such type of authentication is more real because the user has the freedom to record any text. However, in text-dependent speaker recognition, the model authenticates the user with the same text used during the training of the model. The user is required to utter the same text every time for authentication.

In the proposed system, biometric authentication using speech signals is performed in two ways, namely, text-dependent and text-independent speaker identification. Two approaches, one based on MFCC features and the other on PLP features, are applied for authentication. Each approach provides an independent decision about the identity of a user. The experimental results of authentication by speech achieved with the MFCC and PLP are provided in Tables 1 and 2, respectively.

Table 1: Authentication results by speech with MFCC

Speech Signals	Speaker Recognition Type	Training	Testing	Number of Gaussian Mixtures			
				4	8	16	32
Original	Text-dependent	Part 1	Part 1	100%	100%	100%	100%
		Part 2	Part 2	100%	100%	100%	100%
	Text-independent	Part 1	Part 2	92.50%	100%	100%	100%
		Part 2	Part 1	97.50%	100%	100%	100%
Encrypted	Text-dependent	Part 1	Part 1	0%	0%	0%	0%
		Part 2	Part 2	0%	0%	0%	0%
	Text-independent	Part 1	Part 2	0%	0%	0%	0%
		Part 2	Part 1	0%	0%	0%	0%

Table 2: Authentication results by speech with PLP

Speech Signals	Speaker Recognition Type	Training	Testing	Number of Gaussian Mixtures			
				4	8	16	32
Original	Text-dependent	Part 1	Part 1	100%	100%	100%	100%
		Part 2	Part 2	100%	100%	100%	100%
	Text-independent	Part 1	Part 2	92.50	97.50	100%	100%
		Part 2	Part 1	95	97.50	100%	100%
Encrypted	Text-dependent	Part 1	Part 1	0%	0%	0%	0%
		Part 2	Part 2	0%	0%	0%	0%
	Text-independent	Part 1	Part 2	0%	0%	0%	0%
		Part 2	Part 1	0%	0%	0%	0%

The original signals provide the baseline results when no encryption or decryption processes are applied. These results are helpful in making the comparison with the results obtained using encrypted and decrypted signals. Tables 1 and 2 respectively show the results of the original and encrypted signals only because the results of the recovered signals are the same as that of the original signals. Therefore, the results with the decrypted signals are not listed in these tables.

The text-dependent authentication results of MFCC and PLP are 100% with the different numbers of Gaussian mixtures. However, the text-independent speaker authentication with MFCC for four and eight Gaussian mixtures is 92.50% and 97.50%, respectively, when the training is conducted with part 1 of the signal and the testing is performed with part 2 of the signal. For other numbers of Gaussian mixtures, the results are at the maximum. A similar kind of trend can be observed in Table 2, in which all users are authenticated correctly by the the text-dependent system. However, in the case of the text-independent system, a small number of Gaussian mixtures falsely reject some users.

In the encryption process, the encrypted signals are severely distorted, which is also supported by the computed measure EUD. In Tables 1 and 2, the results of authentication with encrypted signals clearly suggest that the encrypted signals do not reveal the identity of a user. Moreover, in the case of data breach, user authentication will not be possible with the encrypted signals. This finding shows that the proposed system is secure and that only a genuine user will be authenticated to access the services.

4.3.2. Authentication Results for Face Recognition

The authentication by face is done for all users in the database. The user cannot send the same image every time. Therefore, in the authentication experiments, the images used to determine the identity of a user are different from those used in the training. The first two images of every user are taken for the authentication, while the remaining eight images are used for the training. The testing images of different users are shown in Fig. 8, which indicates that the testing images vary from user to user. For instance, user 2 has a side movement and different facial expression in the first and second images. User 6 is smiling in the second image but not in the first image. Similarly, the facial expression of user 1 is different in both images and the variation of light can be observed. The first two users are wearing glasses, while the last user is without glasses. Such variation in the testing images is good to observe the robustness of the developed authentication model. The experimental results of authentication by face are presented in Table 3.

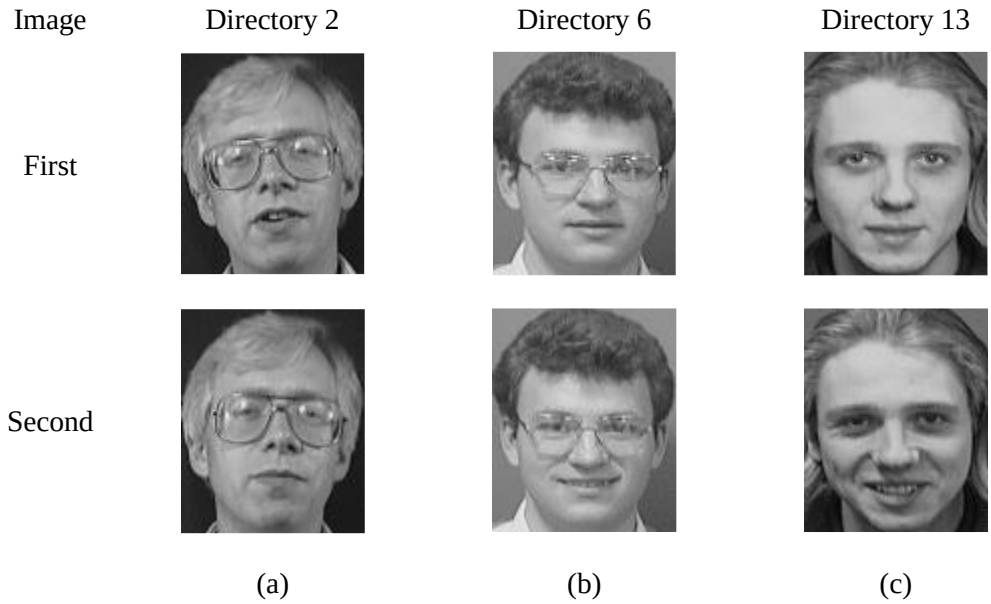


Figure 8: Testing images (first and second images) of different users: (a) side movement/wearing glasses, (b) smiling/not smiling/wearing glasses, and (c) light variation/without glasses.

Table 3: Authentication results by face

Images	Testing Image	
	Image 1	Image 2
Original	97.50%	100%
Encrypted	0%	0%

Table 3 provides the results of the original and encrypted faces. The accuracy of the authentication is 97.50% when the first image of all users is taken for testing. An accuracy of 100% is also achieved when the second image of all users is selected for testing. In Table 3, the authentication using encrypted faces is 0%, indicating that encrypted faces do not reveal the identity, which is good. This finding also supports the conclusion about the reliability of the encryption method deduced in Section 3. The results of authentication using decrypted faces are not presented in Table 3 because the results of the decrypted faces are similar to those of the original faces. The finding that the recovered faces are similar to the original faces shows the reliability of the decryption process.

4.3.3. Combined Authentication Decision

The proposed authentication system provides three decisions about the identity of a person. The speech authentication model delivers two of the decisions, one from each type of feature. The face authentication model provides the third decision. Majority voting makes the final decision about the identity of a user. When any two of the decisions are in favor of a particular user, the proposed system will disclose the identity of that user.

The basic purpose of the authentication models for speech and face is to provide an objective analysis of the proposed encryption method. The results of the models shown in Tables 1, 2, and 3 clearly indicate that

the encryption and decryption of biometrics are accurate and reliable. No one will be allowed to access the services based on the encrypted biometrics. Therefore, the encrypted biometrics are invulnerable and can be transmitted via wireless communication without any threat.

The improvement in the case of multimodal authentication compared with that of the unimodal authentication cannot be observed because each unimodal authentication provided 100% accuracy.

5. Conclusion

A multimodal biometric authentication system is proposed and implemented in this study. The proposed system encrypts the biometric templates of users by applying the new proposed method. Given the protected templates, the proposed system can be used reliably in the centralized cloud environment without fear of information leakage in case of data breach. Moreover, the proposed system uses computing resources in an optimal manner by employing personal portable devices as edges, which ultimately reduce the computational load on the cloud. Experimental results show that the encryption and decryption processes are reliable. Therefore, the identity of the person cannot be disclosed unless it is decrypted with the relevant secret key. The authentication results obtained using encrypted biometrics clearly indicate that the identity of a user cannot be disclosed after the encryption of biometric templates. The proposed system likewise decrypts the biometrics perfectly. Therefore, the results of authentication using the decrypted biometrics are the same as those of the original. In the future, the proposed system will be modified to generate the secret shares of biometric templates to enhance the security of the system.

References

[1] M. Qiu, Z. Chen, Z. Ming, X. Qin, J. Niu, Energy-Aware Data Allocation With Hybrid Memory for Mobile Cloud Systems, IEEE Systems Journal, 11 (2017) 813-822.

[2] M. Qiu, Z. Ming, J. Li, K. Gai, Z. Zong, Phase-Change Memory Optimization for Green Cloud with Genetic Algorithm, IEEE Transactions on Computers, 64 (2015) 3528-3540.

[3] P.G. Lopez, A. Montresor, D. Epema, A. Datta, T. Higashino, A. Iamnitchi, M. Barcellos, P. Felber, E. Riviere, Edge-centric Computing: Vision and Challenges, SIGCOMM Comput. Commun. Rev., 45 (2015) 37-42.

[4] M. Qiu, K. Gai, Z. Xiong, Privacy-preserving wireless communications using bipartite matching in social big data, Future Generation Computer Systems, (2017).

[5] K. Gai, M. Qiu, Z. Ming, H. Zhao, L. Qiu, Spoofing-Jamming Attack Strategy Using Optimal Power Distributions in Wireless Smart Grid Networks, IEEE Transactions on Smart Grid, 8 (2017) 2431-2439.

[6] K. Gai, M. Qiu, Blend Arithmetic Operations on Tensor-based Fully Homomorphic Encryption Over Real Numbers, IEEE Transactions on Industrial Informatics, PP (2017) 1-1.

[7] Z. Ali, G. Muhammad, M.F. Alhamid, An Automatic Health Monitoring System for Patients Suffering From Voice Complications in Smart Cities, IEEE Access, 5 (2017) 3900-3908.

[8] G. Muhammad, M. Alsulaiman, S.U. Amin, A. Ghoneim, M.F. Alhamid, A Facial-Expression Monitoring System for Improved Healthcare in Smart Cities, IEEE Access, 5 (2017) 10871-10881.

[9] P.T. Kim, R.A. Falcone, The use of telemedicine in the care of the pediatric trauma patient, Seminars in Pediatric Surgery, 26 (2017) 47-53.

[10] M. Aguas, J. Del Hoyo, R. Faubel, B. Valdivieso, P. Nos, Telemedicine in the treatment of patients with inflammatory bowel disease, Gastroenterología y Hepatología (English Edition), (2017).

[11] T. Elliott, J. Shih, C. Dinakar, J. Portnoy, S. Fineman, American College of Allergy, Asthma & Immunology Position Paper on the Use of Telemedicine for Allergists, Annals of Allergy, Asthma & Immunology, (2017).

[12] N. Sikka, S. Paradise, M. Shu, Telehealth in emergency medicine: a primer, in, American College of Emergency Physicians, 2014.

[13] O. Ogbanufe, D.J. Kim, Comparing fingerprint-based biometrics authentication versus traditional authentication methods for e-payment, Decision Support Systems, (2017).

[14] W. Dai, M. Qiu, L. Qiu, L. Chen, A. Wu, Who Moved My Data? Privacy Protection in Smartphones, IEEE Communications Magazine, 55 (2017) 20-25.

[15] W. Abdul, Z. Ali, S. Ghouzali, M. Alsulaiman, Security and Privacy for Medical Images Using Chaotic Visual Cryptography, Journal of Medical Imaging and Health Informatics, 7 (2017) 1296-1301.

[16] M.W. Mak, X. Pang, J.T. Chien, Mixture of PLDA for Noise Robust I-Vector Speaker Verification, IEEE/ACM Transactions on Audio, Speech, and Language Processing, 24 (2016) 130-142.

[17] N. Li, M.W. Mak, J.T. Chien, DNN-Driven Mixture of PLDA for Robust Speaker Verification, IEEE/ACM Transactions on Audio, Speech, and Language Processing, 25 (2017) 1371-1383.

[18] H. Li, F. Shen, C. Shen, Y. Yang, Y. Gao, Face recognition using linear representation ensembles, Pattern Recognition, 59 (2016) 72-87.

[19] F. Shen, C. Shen, X. Zhou, Y. Yang, H.T. Shen, Face image classification by pooling raw features, Pattern Recognition, 54 (2016) 94-103.

[20] J. Khodadoust, A.M. Khodadoust, Fingerprint indexing based on minutiae pairs and convex core point, Pattern Recognition, 67 (2017) 110-126.

[21] P. Baldi, Y. Chauvin, Neural Networks for Fingerprint Recognition, Neural Computation, 5 (1993) 402-418.

[22] A.S. Ungureanu, S. Thavalengal, T.E. Cognard, C. Costache, P. Corcoran, Unconstrained palmprint as a smartphone biometric, IEEE Transactions on Consumer Electronics, 63 (2017) 334-342.

[23] G. Li, J. Kim, Palmprint recognition with Local Micro-structure Tetra Pattern, Pattern Recognition, 61 (2017) 29-46.

[24] Y. Zhang, G. Pan, K. Jia, M. Lu, Y. Wang, Z. Wu, Accelerometer-Based Gait Recognition by Sparse Representation of Signature Points With Clusters, IEEE Transactions on Cybernetics, 45 (2015) 1864-1875.

[25] R. Chhatrala, D.V. Jadhav, Multilinear Laplacian discriminant analysis for gait recognition, IET Computer Vision, 11 (2017) 153-160.

[26] B.B. Bhaganagare, A.D. Harale, Iris as biometrics for security system, in: 2017 Second International Conference on Electrical, Computer and Communication Technologies (ICECCT), 2017, pp. 1-7.

[27] K. Nguyen, C. Fookes, R. Jillela, S. Sridharan, A. Ross, Long range iris recognition: A survey, Pattern Recognition, 72 (2017) 123-143.

[28] A. Jain, L. Hong, Y. Kulkarni, A multimodal biometric system using fingerprint, face and speech, in: 2nd International Conference on Audio- and Video-based Biometric Person Authentication, 1999.

[29] A.K. Jain, A. Ross, Learning user-specific parameters in a multibiometric system, in: Proceedings. International Conference on Image Processing, 2002, pp. I-57-I-60 vol.51.

[30] S. Ribaric, I. Fratric, K. Kis, A biometric verification system based on the fusion of palmprint and face features, in: ISPA 2005. Proceedings of the 4th International Symposium on Image and Signal Processing and Analysis, 2005., 2005, pp. 12-17.

[31] F. Besbes, H. Trichili, B. Solaiman, Multimodal Biometric System Based on Fingerprint Identification and Iris Recognition, in: 2008 3rd International Conference on Information and Communication Technologies: From Theory to Applications, 2008, pp. 1-5.

[32] P. Kartik, S.R.M. Prasanna, R.V.S.S.V. Prasad, Multimodal biometric person authentication system using speech and signature features, in: TENCON 2008 - 2008 IEEE Region 10 Conference, 2008, pp. 1-6.

[33] A. Zulfiqar, A. Muhammad, M. M. E. A, A Speaker Identification System Using MFCC Features with VQ Technique, in: 2009 Third International Symposium on Intelligent Information Technology Application, 2009, pp. 115-118.

[34] P. Kartik, R.V.S.S.V. Prasad, S.R.M. Prasanna, Noise robust multimodal biometric person authentication system using face, speech and signature features, in: 2008 Annual IEEE India Conference, 2008, pp. 23-27.

[35] European Parliament, EU Regulation 2016/679 on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation) in, URL http://europa.eu/ rapid/press-release _ IP-15-6321 _ en.htm . 2016.

[36] W. Abdul, Z. Ali, S. Ghouzali, B. Alfawaz, G. Muhammad, M.S. Hossain, Biometric Security Through Visual Encryption for Fog Edge Computing, IEEE Access, 5 (2017) 5531-5538.

[37] M. Gomez-Barrero, E. Maiorana, J. Galbally, P. Campisi, J. Fierrez, Multi-biometric template protection based on Homomorphic Encryption, Pattern Recognition, 67 (2017) 149-163.

[38] M. Gomez-Barrero, E. Maiorana, J. Galbally, P. Campisi, J. Fierrez, Multi-biometric template protection based on Homomorphic Encryption, Pattern Recognition, 67 (2017) 149-163.

[39] F.S. Samaria, A.C. Harter, Parameterisation of a stochastic model for human face identification, in: Proceedings of 1994 IEEE Workshop on Applications of Computer Vision, 1994, pp. 138-142.

[40] Massachusette Eye & Ear Infirmry Voice & Speech LAB, Disordered Voice Database Model 4337 (Ver. 1.03) in, Kay Elemetrics Corp., Lincoln Park, NJ, 1994.

[41] Z. Ali, M. Imran, W. Abdul, M. Shoaib, An Innovative Algorithm for Privacy Protection in a Voice Disorder Detection System, in: A.V. Samsonovich, V.V. Klimov (Eds.) Biologically Inspired Cognitive Architectures (BICA) for Young Scientists: Proceedings of the First International Early Research Career Enhancement School on BICA and Cybersecurity (FIERCES 2017), Springer International Publishing, Cham, 2018, pp. 228-233.

[42] Z. Ali, M. Imran, M. Alsulaiman, T. Zia, M. Shoaib, A zero-watermarking algorithm for privacy protection in biomedical signals, Future Generation Computer Systems, (2017).

[43] Z. Ali, M. Alsulaiman, G. Muhammad, I. Elamvazuthi, A. Al-nasheri, T.A. Mesallam, M. Farahat, K.H. Malki, Intra- and Inter-database Study for Arabic, English, and German Databases: Do Conventional Speech Features Detect Voice Pathology?, Journal of Voice, 31 (2017) 386.e381-386.e388.

[44] H. Pastijn, Chaotic Growth with the Logistic Model of P.-F. Verhulst, in: M. Ausloos, M. Dirickx (Eds.) The Logistic Map and the Route to Chaos: From The Beginnings to Modern Applications, Springer Berlin Heidelberg, Berlin, Heidelberg, 2006, pp. 3-11.

[45] S.H. Strogatz, Nonlinear Dynamics and Chaos, 2nd ed., Westview Press: Boulder, CO, USA, 2014.

[46] H.-O. Peitgen, H. Jürgens, D. Saupe, Chaos and Fractals: New Frontiers of Science, 2nd ed., Springer-Verlag New York, 2004.

[47] Y. Lin, W.H. Abdulla, Principles of Psychoacoustics, in: Audio Watermark: A Comprehensive Foundation Using MATLAB, Springer International Publishing, Cham, 2015, pp. 15-49.

[48] S.Z. Selim, M.A. Ismail, K-Means-Type Algorithms: A Generalized Convergence Theorem and Characterization of Local Optimality, IEEE transactions on pattern analysis and machine intelligence, PAMI-6 (1984) 81-87.

[49] R.A. Redner, H.F. Walker, Mixture Densities, Maximum Likelihood and the EM Algorithm, SIAM Review, 26 (1984) 195-239.

Appendix A

Face Image Directory Number	Name of Speech Signal	Face Image Directory Number	Name of Speech Signal
1	BJB1NRL.wav	21	JKR1NRL.wav
2	AXH1NRL.wav	22	JMC1NRL.wav
3	BJV1NRL.wav	23	JTH1NRL.wav
4	CAD1NRL.wav	24	JXC1NRL.wav
5	CEB1NRL.wav	25	KAN1NRL.wav
6	DAJ1NRL.wav	26	LAD1NRL.wav
7	DFP1NRL.wav	27	LDP1NRL.wav
8	DJG1NRL.wav	28	LLA1NRL.wav
9	DMA1NRL.wav	29	LMV1NRL.wav
10	DWS1NRL.wav	30	LMW1NRL.wav
11	EDC1NRL.wav	31	MAM1NRL.wav
12	EJC1NRL.wav	32	MAS1NRL.wav
13	FMB1NRL.wav	33	MCB1NRL.wav
14	GPC1NRL.wav	34	MFM1NRL.wav
15	GZZ1NRL.wav	35	MJU1NRL.wav
16	HBL1NRL.wav	36	MXB1NRL.wav
17	JAF1NRL.wav	37	MXZ1NRL.wav
18	JAN1NRL.wav	38	NJS1NRL.wav
19	JAP1NRL.wav	39	OVK1NRL.wav
20	JEG1NRL.wav	40	PBD1NRL.wav